

Information and Modeling Issues in Designing Water and Sanitation Subsidy Schemes

Andrés Gómez-Lobo

Vivien Foster

Jonathan Halpern

Evaluating design alternatives is a first step in introducing optimal water subsidy schemes. The definition of appropriate targeting criteria and subsidy levels needs to be supported by empirical analysis, generally an informationally demanding exercise. An assessment carried out in Panama revealed that targeting individual households would be preferable to geographically based targeting. Empirical analysis also showed that only a small group of very poor households needed a subsidy to pay their water bill.

The World Bank
Latin America and the Caribbean Region
Finance, Private Sector, and Infrastructure Sector Unit
May 2000



Summary findings

In designing a rational scheme for subsidizing water services, it is important to support the choice of design parameters with empirical analysis that simulates the impact of subsidy options on the target population. Otherwise, there is little guarantee that the subsidy program will meet its objectives.

But such analysis is informationally demanding. Ideally, researchers should have access to a single, consistent data set containing household-level information on consumption, willingness to pay, and a range of socioeconomic characteristics. Such a comprehensive data set will rarely exist. Gómez-Lobo, Foster, and Halpern suggest overcoming this data deficiency by collating and imaginatively manipulating different sources of data to generate estimates of the missing variables.

The most valuable sources of information, they explain, are likely to be the following:

- Customer databases of the water company, which provide robust information on the measured consumption of formal customers but little information on unmeasured consumption, informal customers, willingness to pay, or socioeconomic variables.
- General socioeconomic household surveys, which are an excellent source of socioeconomic information but tend to record water expenditure rather than physical consumption.

- Willingness-to-pay surveys, which are generally tailored to a specific project, are very flexible, and may be the only source of willingness-to-pay data. However, they are expensive to undertake and the information collected is based on hypothetical rather than real behavior. Where such surveys are unavailable, international benchmark values on willingness to pay may be used.

Combining data sets requires some effort and creativity, and creates difficulties of its own. But once a suitable data set has been constructed, a simulation model can be created using simple spreadsheet software. The model used to design Panama's water subsidy proposal addressed these questions:

- What are the targeting properties of different eligibility criteria for the subsidy?
- How large should the subsidy be?
- How much will the subsidy scheme cost, including administrative costs?

Armed with the above information, policymakers should be in a position to design a subsidy program that reaches the intended beneficiaries, provides them with the level of financial support that is strictly necessary, meets the overall budget restrictions, and does not waste an excessive amount of funding on administrative costs.

This paper — a product of the Finance, Private Sector, and Infrastructure Sector Unit, Latin America and the Caribbean Region — is part of a larger effort in the region to evaluate and disseminate lessons of experience in designing policies to improve the quality and sustainability of infrastructure services and to enhance the access of the poor to these basic services. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Silvia Delgado, room IS-196, telephone 202-473-7840, fax 202-676-1821, email address sdelgado@worldbank.org. Policy Research Working Papers are also posted on the Web at www.worldbank.org/research/workingpapers. The authors may be contacted at vfoster@worldbank.org or jhalpern@worldbank.org. May 2000. (35 pages)

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.

**INFORMATION AND MODELING ISSUES
IN DESIGNING WATER AND SANITATION
SUBSIDY SCHEMES**

**Andrés Gómez-Lobo¹
Vivien Foster²
Jonathan Halpern³**

¹ Universidad de Chile.

² LCSPR, Latin America and the Caribbean Region, the World Bank Group.

³ LCSFP, Latin America and the Caribbean Region, the World Bank Group.

TABLE OF CONTENTS

	<u>Page</u>
TABLE OF CONTENTS.....	2
EXECUTIVE SUMMARY	I
1. INTRODUCTION.....	1
2. INFORMATIONAL REQUIREMENTS	2
<i>Willingness-to-Pay Data</i>	2
<i>Water Consumption Data</i>	3
<i>Socio-Economic Data</i>	4
3. SOURCES OF INFORMATION.....	5
<i>Client Databases</i>	5
<i>Household Surveys</i>	6
<i>Census Data</i>	9
<i>Willingness to Pay Data and Surveys</i>	10
<i>Other Data Sources</i>	14
4. DEALING WITH INCOMPLETE DATA SETS.....	16
<i>Inferring Unmeasured Consumption</i>	16
<i>Inferring Consumption from Expenditure Data</i>	18
<i>Inferring Consumption from Socio-Economic Data</i>	20
5. MODELLING AND SIMULATION ISSUES.....	24
<i>Determining Eligibility Criteria</i>	24
<i>Establishing the Magnitude of Subsidy</i>	26
<i>Estimating Budgetary Requirements</i>	28
6. CONCLUSIONS	32
REFERENCES.....	34

EXECUTIVE SUMMARY

This paper discusses some of the practical issues that have to be dealt with when designing a rational water subsidy scheme. A central conclusion is that the choice of suitable design parameters for a subsidy scheme needs to be supported by appropriate empirical analysis to simulate the impact of alternative types of subsidies on the ultimate target population. Without such investigation, there is little guarantee that a subsidy system, however well meant, will meet its intended objectives.

Nevertheless, it is acknowledged that the kind of analysis required to design a subsidy programme is relatively demanding in terms of information. Ideally, the policymaker would need to have access to a single and consistent data set containing information on willingness to pay, water consumption and a range of socio-economic characteristics, each reported at the household level.

Clearly, such comprehensive information will rarely be available in practice. However, this in itself should not be regarded as a sufficient reason to abandon the counsels given above. As is illustrated in the paper, deficiencies in the available information can to some extent be overcome by collating different sources of data and imaginatively manipulating this data so as to generate estimates of the missing variables.

Based on the review contained in the paper, the most valuable sources of information are likely to be the following.

- *Water company client databases*—these provide robust information on the measured consumption of formal customers, but provide little information about unmeasured consumption or informal customers. The main drawback with this data source is that it will not contain any information on the socio-economic characteristics of the clients or their willingness to pay.
- *General socio-economic household surveys*—most countries conduct extensive household surveys on an occasional basis, many of them modelled on the World Bank's 'Living Standards Measurement Survey' blueprint. Such surveys provide excellent socio-economic data, but tend to record water expenditure rather than water consumption. Various deficiencies in the recording of water expenditure can make it difficult to make reliable inferences about water consumption.
- *Willingness-to-pay surveys*—these are generally tailor-made surveys conducted in the context of a specific project. They are an extremely flexible tool for determining willingness to pay, and may also be used to complete other gaps in the information base. However, they are subject to a range of methodological criticisms. Perhaps the most fundamental of these is the fact that they are based on hypothetical as opposed to real behavior. Where it is not possible to conduct a willingness-to-pay survey, international benchmark values may be used, based on previous experience.

Once a suitable data set has been constructed from these sources, a simulation model can be created using simple spreadsheet software. The model should be capable of addressing the following key design questions.

- *What should the eligibility criteria be?* Subsidies are generally assigned on the basis of eligibility criteria presumed to have a strong correlation with poverty. Using the subsidy data set, the targeting performance of alternative eligibility criteria can be compared by examining the proportion of the poor who would meet any particular eligibility criterion. It is also possible to examine what proportion of the eligible population are not genuinely poor, for any particular set of criteria.
- *How large should the subsidy be?* Economic theory suggests that the subsidy should be set equal to the difference between the willingness to pay of the poor and the cost of purchasing a subsistence level of consumption. However, willingness to pay is likely to vary with household income and subsistence consumption with household composition. Using the simulation model, it is possible to experiment with alternative reference levels of income and consumption, and to examine what proportion of poor households would be adequately protected by any given level of the subsidy.
- *How much will the subsidy scheme cost?* The costs of the subsidy programme must necessarily be commensurate with the available funding. A simulation model allows ready estimation of the overall costs of the subsidy programme. These depend not only on the size and scope of the subsidy, but also on the rate of uptake among eligible households, as well as the rate of infiltration of non-eligible households. The administrative costs of the programme also need to be taken into account, and, in particular, the proportion of total costs that are absorbed by administrative procedures.

Armed with this information, the policymaker should be in a position to design a subsidy programme that reaches the intended beneficiaries, provides them with the level of financial support which is strictly necessary, meets the overall budgetary restrictions, and does not waste an excessive amount of funding on administrative costs.

This type of approach was used in the design a subsidy scheme for the water sector in Panama. Although the proposed subsidy has not been implemented yet, the modelling approach was useful to discard the use—due to its poor targeting properties—of a geographically based subsidy scheme in favor of an individually targeted one. In addition, the careful empirical analysis conducted in that study showed that most households did not require a subsidy in order to afford their water consumption bill. This evidence provided a justification for keeping the water subsidy programme focused on a relatively small group of very poor households. It was also instrumental in refocusing the proposed scheme towards a subsidy for sewerage installation, where a big disparity between bills and willingness to pay was identified.

1. INTRODUCTION

In the developing countries, water as a commodity is often subsidized. These subsidies are the result of a general underpricing of water, numerous cross-subsidies and the lack of efficient billing and collection. With few exceptions, there are generally no explicit welfare policy objectives behind these subsidies. They are the result of historical and political events, or simply—as in the case of inadequate billing and revenue collection—the unintended (or perhaps intended) side effect of managerial inefficiency.

One of the stated aims of sectoral reform processes is the rationalization of water tariffs and subsidies. Ideally, policymakers must decide who they wish to subsidize and by how much, and select the preferred instrument for reaching these policy objectives. In other words, explicit subsidies, with well-defined objectives, budgets and instruments, should be designed to replace the existing implicit subsidy arrangements.

The objective of this technical note is to describe the informational and modelling issues that arise when designing a direct subsidy scheme for the water sector. There is no discussion of the rationale for these subsidies, nor of all the policy options available to the authorities, since these issues are already addressed in the accompanying paper entitled, 'Design of Direct Subsidy Systems for Water and Sanitation Services: A Case Study from Panama'. Rather, the present paper seeks to complement the policy discussion by providing a thorough treatment of the technical challenges that arise in seeking to implement the subsidy design methodology expounded in the companion paper.

Specifically, this paper will deal with the following issues.

- Section 2 identifies the different types of data that are required in order to make informed judgements about subsidy design options.
- Section 3 reviews the main sources of information that are available to the policy analyst, and compares their relative strengths and weaknesses.
- Section 4 provides a number of practical examples of how to overcome deficiencies in data sources by creative inference of key variables.
- Section 5 describes how the information collected can be used to simulate the effects of alternative subsidy design parameters using simple modelling techniques.
- Section 6 draws out the main lessons and conclusions from the rest of the paper.

As in the companion paper, the discussion will draw extensively on illustrative case material derived from a water subsidy design exercise undertaken for the Republic of Panama in 1998.⁴

⁴ The study was undertaken by Oxford Economic Research Associates Ltd., in association with ESA Consultores of Honduras, and EMG Consultores of Chile. Part of the credit for the work presented here is due to them.

2. INFORMATIONAL REQUIREMENTS

In order to develop an optimal subsidy scheme, there are numerous questions that need to be addressed. Of these, many require a quantitative answer. Important examples are the following.

- What are the targeting properties of alternative eligibility criteria?
- How large should the subsidy be for each eligible user?
- What are the budgetary requirements for any particular subsidy system?

To answer these questions, three types of data must be collected:

- willingness-to-pay data;
- water consumption data;
- socio-economic data.

The purpose for collecting each type of information will be discussed in turn.

Willingness-to-Pay Data

Ideally, an analyst would like to have information on users' willingness to pay for water and sanitation services. Willingness to pay is the maximum amount that a household would be prepared to spend to secure access to a given quantity of the service. Thus, in economic terms, it represents the limit of affordability of the service. Strictly speaking, it is equivalent to the area under the demand function for any particular consumption level.

Unless political or other considerations are to be taken into account, a reasonable rule would be to set subsidy levels to cover the shortfall between a vulnerable household's willingness to pay for a basic level of consumption and the associated bill. In formal terms, this idea can be expressed as follows.

$$S = \begin{cases} T(c) - WTP(c) & \text{if } T(c) - WTP(c) > 0 \\ 0 & \text{if } T(c) - WTP(c) \leq 0 \end{cases}$$

Where, S is the level of subsidy accruing to each beneficiary household, $T(c)$ is the tariff associated with the *subsistence* level of consumption, c , and $WTP(c)$ is the willingness to pay for the service. It is important to note that subsidies should not be applied beyond the basic subsistence level of consumption, in order to avoid encouraging excessive consumption or diluting incentives to detect and repair leaks within the dwelling. Furthermore, such subsidies should be confined to those sectors of the population that are genuinely vulnerable.

Note that an analogous formula can be developed for connections or any other service that may be the subject of a subsidy.

The justification for the above rule is the following. In principle, the aims of a water subsidy scheme are to promote the connection of vulnerable households (and avoid their subsequent disconnection), and to ensure that, once connected, they attain a level of consumption compatible with the attainment of public health objectives. If the amount of the subsidy is set according to the above rule—and provided the WTP data truly reflects users' preferences and future consumption decisions—then the subsidy will be sufficient to induce the user to consume at least the socially desirable minimum amount of water.

Of course, governments may have political or distributional motivations for introducing a water subsidy scheme, which may be unrelated or additional to the strict “affordability” criteria set out above. However, the working assumption of this paper is that the justification of a rational water subsidy scheme is to induce users to consume the socially desirable level of services. In part, this is due to the fact that governments generally have alternative instruments to distribute income that are better suited to that task than a sectoral consumption subsidy.

Water Consumption Data

Another important piece of information is the pattern of demand for different types of households. This data is valuable for several reasons.

- First, it can be used to establish the basic consumption level that will be subsidized, or, in other words, to determine what should be regarded as a subsistence consumption level in any particular instance. A good way of doing this is to examine the frequency distribution of consumption for vulnerable households. A cut-off point can then be selected (for example, the average consumption of vulnerable households or of the poorest 25% of these households). Alternatively, other sources of information could be used to derive this basic or ‘subsistence’ consumption level, including regional experience or engineering estimates of household basic usage. However, even in these cases, it remains advisable to check this external benchmark against real consumption data to make certain that it is not grossly under- or overestimating actual requirements.
- Second, it may be desirable for the subsidy to reflect the seasonal patterns of water consumption. For example, if basic or ‘subsistence’ consumption is higher during certain months of the year, the subsidy may need to reflect this. Although a subsidy calculated on a yearly basis will, on average, give the same benefit to users as a seasonal differentiated subsidy, the existence of liquidity constraints on the part of households may make it desirable to have varying degrees of subsidies during the year. This is especially so if tariffs have a peak pricing component that substantially increases bills during the peak consumption period.
- Third, where users are not currently metered (so that the above two points become somewhat irrelevant), it still remains useful to have information on water consumption patterns, particularly where there is a policy to increase metering. It is convenient to check that the minimum consumption level set for the subsidy is

consistent with the real consumption of those households which are expected to be metered in the future. This will reduce the need to readjust the subsidy as the metering programme progresses. This requires estimating the consumption level of these non-metered households. As a by-product, knowledge of non-metered consumption can be very useful to estimate the water company's levels of physical and commercial losses.

Socio-Economic Data

The final piece of information required is socio-economic data. These should ideally include household income or expenditure levels, poverty lines, and receipt of other welfare benefits, as well as general indicators of basic needs (such as the quality of housing and its associated facilities), and family wealth (such as durable goods ownership and the property value of the dwelling). This data is useful for the following purposes.

- To determine the target population group for the subsidy (i.e., whom is the subsidy attempting to benefit?)
- To study the targeting properties of different eligibility criteria used to distribute the subsidy (such as geographic location, wealth indicators or other variables).
- To determine the proportion of household income that is spent on water and sanitation. Often willingness-to-pay parameters are expressed as a percentage of household income or expenditure. In order to compare this threshold with the household water bill, it is necessary to have a measure of income or expenditure.

3. SOURCES OF INFORMATION

Ideally, it would be desirable to have a data set that contains simultaneously all of the above information for each household (h). With consumption (c), willingness to pay (wtp) and socio-economic characteristics (Z) recorded for each observation in the sample, it becomes straightforward to simulate the impact of alternative subsidy designs.

$$h_i(c, wtp, Z)$$

A full-scale demand study would be capable of generating all the necessary information. This is because the willingness to pay for a certain amount of a good is given by the area under the demand curve, while consumption behavior is also determined by the demand function. Furthermore, if exogenous socio-economic factors are included as determinants of water demand, and the distributions of these factors in the population are known, then all the required information would be available from one source.

However, in most cases, a comprehensive demand study of this kind will not be available. Even where such a study exists, it is unlikely that it will include all the socio-economic variables that are relevant for the design of a water subsidy. For example, geographic location or property valuations are most probably not important determinants of demand. However, for the targeting of a subsidy, they are crucial variables to analyze for their potential use as eligibility criteria. In addition, it is unlikely that a demand study will include a very detailed analysis of a household's total income or expenditure, which requires much effort to estimate correctly.

Most often, the ideal database will not be available, so that the required information will have to be collated from different sources. Some degree of flexibility and creativity will therefore be required in order to answer the questions posed at the beginning of this paper. The main sources of relevant information are the following:

- client databases;
- household surveys;
- census data;
- willingness-to-pay surveys.

Client Databases

A first source of information is the company's client database. This will typically contain monthly information on the number of customers, the value of their bills, the level of consumption for measured customers, the location of customers, and the tariff structure applicable to each customer. Thus, client databases are generally the best source of information on consumption levels (the c variable defined above).

The client database provides a helpful starting point, in particular because it can be used to create a frequency distribution of consumption patterns for measured customers, and to determine seasonal trends in consumption.

For example, under Chilean law, the government is allowed to subsidize water consumption *up to a maximum threshold* of 20m^3 per family per month in the case of vulnerable households. In order to establish exactly what the subsistence consumption level should be, the Ministry of Planning studied the client databases of the various regional water companies in order to understand water consumption patterns. This exercise was facilitated by the fact that there is universal metering of water in Chile. The analysis revealed that 80% of households consumed less than 20m^3 per month. This finding suggested that the maximum 20m^3 threshold was too generous an allowance for subsistence consumption, so that the subsistence consumption level to be used for the purposes of the subsidy was reduced to 15m^3 per household per month.

However, there are several drawbacks of a client database as a source of information.

- A client database will not include information on informal connections. As part of the reform process, there may be a campaign to formalize such users. The subsidy scheme may be designed to encourage such a process. If so, then the client database may be of limited use for the purposes of estimating the required individual subsidy, as well as estimating the total budget required for the subsidy programme in the future.
- A client database is not very informative as to the consumption level of unmeasured clients. However, a method for estimating this information indirectly from the client database will be presented later in the paper.
- The most important drawback of a client database as a source of information is the lack of data on the socio-economic characteristics of each household. Although the water company may have a special tariff for vulnerable households, rarely will the beneficiaries of this tariff correspond to a strict and objective definition of a low-income or disadvantaged household. For example, in the case of Panama, about two-thirds of the customer base benefit from concessional tariffs, whereas only 16% of the customer base live in conditions of poverty according to official definitions. For this reason alone, the client database rarely suffices as a source of information for subsidy design.

In Panama, a data set was created using the client records of the water utility, IDAAN. This database contained information on the location of clients, the corresponding tariff structure, and the consumption charged each month (whether measured, estimated or imputed).

Household Surveys

A very useful source of information is the data from household level expenditure or socio-economic surveys. These surveys are routinely undertaken by the statistical offices of most countries in order to update consumer price indices, to measure poverty levels or to obtain other socio-economic information. Household surveys are probably the best source of

information to ascertain poverty levels, demographic characteristics, and general socio-economic information (the Z vector defined above).

Many countries follow some form of the Living Standards Measurement Study survey methodology, which was developed by the World Bank in 1980 and has subsequently been adopted (with minor variations) in more than 20 developing countries, including Panama. These surveys are one of the most popular tools for measuring living standards and poverty, as well as designing government policies and evaluating social programmes.⁵ The detailed socio-economic measurements contained in these surveys make them an important source of information for designing a water subsidy scheme.

A suitable database of this kind can provide the following invaluable information.

- Estimates of family weekly, monthly or yearly income (or expenditure), that can then be divided by some measure of family composition to obtain an equivalent income (or expenditure) figure⁶. These figures can then be used to ascertain the number of households that fall below a given poverty line. Such poverty lines are usually established by the official statistical agency in any particular country, often based on information collected in the Living Standards Measurement Survey.
- Ownership of durable goods, dwelling characteristics, access to other utility services and other socio-economic data provide helpful alternative indicators of poverty within the context of a basic needs approach. This information is particularly important for examining the targeting properties of alternative subsidy eligibility rules, which tend to be based on readily observable household characteristics, rather than self-reported household income.

Although household surveys are extremely useful for a subsidy design exercise, they also carry a number of general drawbacks.

- These surveys are expensive and time-consuming to conduct, and governments therefore do not usually undertake them often, perhaps once every four or five years. This implies that, in some countries, the information may be somewhat outdated as compared to the client database, for example.
- Surveys are representative of the population according to a pre-defined sampling design frame (for example, the urban and rural areas of a country). They may not be representative of the population in the areas or zones relevant for the subsidy design (say, province or city level).

⁵ Deaton (1997) presents an in-depth analysis of the potential uses of these data sets.

⁶ If the family composition index is just the headcount of household members, then the resulting equivalent income is the income per capita of the household. However, economists sometimes weight children and adults, as well as adults besides the head of household, differently to reflect the economies of scale in household production.

- Related to the above point, household surveys will not be very useful for classifying households based on fine spatial or geographic location. For example, an analyst would like to examine the properties of a subsidy targeted according to the specific location of a user (e.g., a neighborhood or block in a city). Surveys will usually not include this information, nor will they in general be representative at this level of disaggregation.

Beyond the socio-economic information they contain, most living standards measurement surveys include questions regarding water supply and sanitation in the housing module of the questionnaire.⁷ This makes them particularly valuable in that they are potentially the only source of data that combines information on water use and wider socio-economic characteristics for particular households.

Typical questions include the source of water supply, the average number of hours a day in which a dwelling receives water, and whether there is a sewerage connection. Other questions that are sometimes included are the distance of the dwelling to the water supply, location of the tap, and other characteristics of the water and sewerage services.

As regards water usage, all surveys incorporate a question on the amount the household spent on water services during the last month or the last payment period, although they do not tend to record the volume of water consumed. As such, the only way that water use can be inferred from the information collected in the Living Standards Measurement Survey is to transform the monetary expenditure into a physical consumption variable by applying the corresponding tariff structure to the household's declared water bill.

Experience with applying this approach in Panama, see Section 4.3 below, revealed that the expenditure information was deficient in a number of respects, which made it very difficult to draw reliable inferences about the physical volume of consumption. In particular, these include the following problems.

- The fact that there are multiple tariff structures applied to residential customers and that the survey did not contain any information on which tariff applies to which household.
- The absence of a variable identifying whether the household has measured water supply. Therefore, it is impossible to know whether the expenditure transformation gives actual or imputed water consumption.
- The quality of the expenditure data can be poor. Where the household was not able to produce a recent water bill, the estimate is based on memory. In these cases, it is not always clear whether the estimated consumption includes the charge for refuse collection, which is billed together with the water service.

⁷ The discussion that follows is based on an analysis of the questionnaires for the following surveys: Panama (1997), Peru (1994), Albania (1996), Bulgaria (1995), Ecuador (1995), Kyrgyzstan (1993), and Nepal (1996).

Based on the Panama experience, several recommendations can be made to improve the usefulness of the water sector information of Living Standards Measurement Surveys and other related survey instruments.

- The interviewer training should emphasize the importance of asking for the actual bill in relation to expenditure on publicly supplied water and sewerage services.
- Where the bill is not available, the interviewer should remind the interviewee to exclude from the estimate the costs of irrelevant services such as refuse collection.
- Furthermore, in these situations, the interviewer should register the fact that the expenditure recorded is based on the household's recollection rather than on the bill itself.
- A quality-control procedure should be implemented, whereby the interviewer alerts the household when a stated expenditure figure (based on recollection) is implausible (for example, if it is below the minimum charge of the tariff structure).
- The survey should record whether the household has a water meter installed.

If the interviewer is able to see the actual water bill, it would be desirable to annotate the following information.⁸

- Whether the last bill was based on measured, estimated (based on past meter readings), or imputed consumption.
- The particular tariff that was applied to the household (if this information is clearly specified in the bill).
- The amount of water consumed in physical units (if this information is recorded in the bill).

The recommendations presented are relatively simple to undertake and would not increase the burden or costs of administering a survey. All that is required is to train interviewers to identify the above information from actual water bills, which, under current practice, they are obliged to see anyway.

Census Data

Another useful source of information on socio-economic characteristics may be the national census of a country. Census information usually includes a wide range of demographic, durable ownership and other socio-economic variables. They have several advantages over household surveys.

- By definition, censuses are representative at all levels of disaggregation, including very fine geographic locations.

⁸ A more ambitious strategy would be to record the customer number of the client. This variable could then be used to cross the household survey data with the Water Company's client database. However, unless the statistical office did the crossing of data sets itself, before the information is made public, this alternative would probably violate the confidentiality rules that apply to such surveys.

- Related to the above, they can be used (and are used by official bodies in many countries) to generate poverty maps, which identify the neighborhoods or areas where the poor and vulnerable households are concentrated. This may be very useful information if geographic subsidies are to be considered.

However, census data has important drawbacks that limit their use for the design of a water subsidy scheme.

- The problems of outdated information may be more severe than in the case of surveys. As a general rule, censuses are undertaken once every ten years or so in developing countries.
- Another problem with census data is that income or expenditure are usually (if at all) poorly measured. This is a limitation if poverty is going to be assessed based on an equivalent income or expenditure basis.
- The information regarding water consumption and bills will be at least as bad (if not worse) than in the case of household surveys.

In Panama, census data had been used by a number of government institutions to create geographically based 'poverty maps'. Unfortunately, these poverty maps proved to be of limited value in the design of the water subsidy scheme. This was partly because they were eight years out of date, but also because they were based on relatively large geographic zones and did not contain very reliable information on household income.

Willingness to Pay Data and Surveys

Willingness to pay information is probably the most difficult to obtain. There are four potential sources for this information:

- econometric demand estimations;
- surrogate markets;
- specially designed surveys;
- international comparisons and benchmarks.

As noted above, willingness to pay corresponds to the area underneath the demand curve. Econometric studies can be undertaken to recover the demand function based on past consumption behavior.⁹ However, the applicability of this tool will probably be very limited for the following reasons.

⁹ See Hansen (1996), Hewitt and Hanemann (1995), Lyman (1992), Nieswiadomy (1992), Nieswiadomy and Molina (1989), Stephens, Miller and Willis (1992) for studies that estimate water demand functions.

- Estimating demand functions is not always easy. Besides the statistical and econometric problems faced, there is the added difficulty of obtaining a suitable data set that not only includes consumption by household and socio-economic variables, but also sufficient price variations to identify demand reactions;
- If consumers are not measured then they do not face a marginal price, and it would be impossible to estimate a demand function from past water consumption behavior.
- Difficulties in inferring willingness to pay from past demand behavior may be compounded when there are differences in the service quality received by different users.¹⁰ For example, poor households (mostly unmeasured) may not receive a continuous service. Estimating their willingness to pay based on the behavior of richer (and measured) households, who receive a continuous service may be misleading.

If a full-scale econometric demand study is not available, a second possibility is to use a surrogate market approach. In this case, willingness to pay is inferred based on the behavior of households in a market for a good that is a complement or substitute for public water supply.¹¹ The obvious example is demand for non-piped sources of water (such as bottled water or truck-based vendors).

For example, this method was used in the Dominican Republic to provide a crude first-cut estimate of the demand function for piped water services. Information obtained from local sources indicated that households which lacked a connection to the public network were typically obtaining water at a cost of US\$6.33/m³ from water tankers, and were tending to consume approximately 7m³ per household per month. That is equivalent to a total expenditure of US\$44.3 per month. Whereas households with a connection to the public network were obtaining water at around US\$0.13/m³ and were consuming on average 37m³ per household per month. That is equivalent to a total expenditure of US\$4.93 per month. These two situations essentially describe two points on the demand curve for water, which can be used to make inferences about the overall demand curve and the corresponding willingness to pay.¹² However, the surrogate market approach also has its limitations.

- It is only applicable for households that are not connected or that have very infrequent service, which may not be the case for most of the target population of the survey.
- It is highly questionable whether the data about consumption of piped and non-piped water really represent two points on the same demand curve. First, the quality of the

¹⁰ See Carpentier and Vermersch (1997) (and the references contained therein) for an estimation strategy of the willingness to pay for different qualities of potable water.

¹¹ Both the surrogate market approach and the contingent valuation methodology discussed next are techniques that have been extensively developed in the environmental economics literature. A more rigorous and complete treatment of these and other techniques can be found in Freeman III (1994).

¹² In order to do this, it is necessary to make a (more or less arbitrary) assumption about the functional form of the demand curve. Two commonly used functional forms are the linear and the iso-elastic specifications.

two services is significantly different. For another, the socio-economic characteristics of the two groups of consumers may be very divergent.

A third alternative is to measure willingness to pay is to undertake a specific willingness-to-pay survey. These surveys use a 'contingent valuation' approach to deduce users' preferences for the service. This approach is based on the construction of a hypothetical scenario in which the respondents are asked to consider their willingness to pay for a particular change in the quality or quantity of the service. There are several advantages to this method.

- Since it based on a questionnaire designed by the researcher, it is very flexible, allowing for the estimation of the willingness to pay for a wide range of services and quality levels (e.g., connections to water, connection to sewerage, water consumption, treatment, etc.)
- It is relatively quick, although not inexpensive. A willingness-to-pay survey can be undertaken within a few months, but may well cost in the order of US\$100,000.
- The survey may be used to obtain a useful set of complementary socio-economic information, including the ownership of durable goods, the dwelling conditions, whether the household is metered, the imputed or actual metered consumption (when the interviewer can see the household's water bill), and other relevant variables.
- As a by-product, willingness-to-pay surveys can also generate useful information about consumers' attitudes on a range of subjects relating to the water sector, including their attitudes towards different types of subsidies.

Willingness-to-pay surveys are not, however, without their drawbacks.

- The willingness-to-pay figures are not based on actual payment behavior, but rather on the respondents' replies to hypothetical questions. If respondents do not really believe that they would ever have to pay the amount they state in the survey, they may overstate their willingness to pay, for a number of reasons. For example, they may not have given enough thought to the other claims on their income, they may be embarrassed to admit to the interviewer that they would not really be willing to pay, or there may be strategic reasons why they may want to influence the ultimate policy decision. Indeed, there is a significant body of evidence from laboratory experiments to suggest that hypothetical willingness to pay can substantially exceed the real payments that consumers are prepared to make in comparable circumstances.¹³ As a result, it has sometimes even been recommended that estimates obtained from willingness-to-pay surveys should be divided in half before being applied to inform policy choices (National Oceanographic and Atmospheric Administration, 1994).

¹³ For a review of the experimental literature on the divergence between real and hypothetical willingness to pay, see Foster, V. *et al.* (1997).

- Numerous empirical studies have shown that the results obtained from contingent valuation surveys are highly sensitive to the context and manner in which the willingness-to-pay question is posed.¹⁴ In particular, results have been found to vary substantially depending on the following issues. First, the method used to elicit willingness to pay can have a major effect; in particular whether this is through an open-ended or closed question.¹⁵ Second, responses are sensitive to the amount of information that is supplied about the good or service in question. Third, respondents' attitudes towards the institution that would be receiving the payment may colour the responses obtained (for example, whether the money would be paid to government, a private company or a charitable organization). Fourth, the order in which willingness-to-pay questions are asked, and the nature of the material that immediately precedes them in the questionnaire, are also pertinent factors. All of these choices about questionnaire design can introduce potential biases into the values obtained from survey respondents.
- After the data have been collected, the generation of willingness-to-pay estimates typically require a further stage of statistical analysis.¹⁶ Here, once again, potential distortions can creep in, since a number of comparatively subjective decisions about the empirical analysis can have a substantial effect on the ultimate estimate of willingness to pay. These include the treatment of non-responses, or suspiciously large outlier values, as well as the choice of functional form for econometric estimation.
- From the point of view of subsidy design, a further disadvantage of these surveys is that they may not record household income or expenditure very accurately. To measure these variables, surveys have to be specifically designed and are usually much more extensive and expensive than contingent valuation surveys.

Notwithstanding these problems, numerous willingness-to-pay surveys have been conducted for the water sector. In Latin America, for example, such studies have been carried out in Caracas, Barquisimeto and Merida (Venezuela), Tegucigalpa (Honduras), Managua (Nicaragua) and Guatemala City (Guatemala).¹⁷ Some results of these studies are presented in Table 3.1, among them the outcome of the study conducted in Panama.

¹⁴ For an extensive discussion of the pitfalls of contingent valuation, the reader is referred to the seminal paper by Diamond and Hausman (1994).

¹⁵ An open-ended question is one which takes the form, 'How much are you willing to pay?' while a closed question is one which takes the form, 'Are you willing to pay X?', followed by a 'Yes' or 'No' answer.

¹⁶ Ardila *et al.* (1998) comment on the distortions that can arise at the econometric stage. They also provide a wider panorama of the use of contingent valuation studies in project appraisals by multi-lateral institutions.

¹⁷ These studies were all conducted by ESA Consultores, Tegucigalpa, Honduras.

Table 3.1 Results of willingness-to-pay surveys for water supply in Latin America

City	Year	Results
Tegucigalpa	1995	Marginal households with poor service were on average willing to pay up to 3% of income for improvement
Guatemala City	1997	Marginal households with poor service were on average willing to pay up to 3.5% of income for improvement. Those without access to piped service were on average willing to pay up to 5% of income for the service.
Managua	1995	Connection charges decrease substantially the probability of connection.
Panama	1998	Households in the lowest three deciles of the income distribution were on average willing to pay 4% of their household income for a good water service.
Caracas	1997	Marginal households with bad service were on average willing to pay up to 2.6% of income for improvement. Connection charges decrease substantially the probability of connection.

Source: ESA Consultores.

If a willingness-to-pay survey is not undertaken, the final option is to use parameters taken from other regional or international studies of willingness to pay, or to use a benchmark parameter. Thus, the values of Table 3.1 could be used to determine a range of feasible willingness-to-pay values (expressed as a percentage of family income or expenditure). It is striking that the results from this range of studies are comparatively consistent with the willingness to pay of poorer households for piped water supplies invariably falling in the range 3–5% of household income.

A similar approach is to make use of some standard benchmark parameter. A popular reference value that has been adopted in some countries is that prescribed by the Pan American Health Organization, which recommends that household water bills should not represent more than 5% of household income (3.5% for water, 1.5% for sewerage). In Chile, for example, an explicit target of the water subsidy scheme is to reduce households' water bills below this 5% benchmark. The advantage of using parameters from international or regional experiences is that they are readily available. However, there is also the disadvantage that the particular parameters chosen may not be applicable to the particular area under study.

Other Data Sources

There are a number of other data sources that may provide relevant information for the design of a water subsidy scheme. These include the following.

- Property value registers can be of interest, in so far as property value can be regarded as a reasonable proxy for household wealth, and therefore as a potential eligibility criterion for receipt of the subsidy. In practice, many property value registers can be very incomplete and often highly out of date, factors that limit their value in this context.

- Databases kept by other public entities that are providing welfare benefits—this information may help to identify vulnerable households.

Most of these other data sources will contain information on a limited set of variables, and the question arises as to how to integrate these variables with the information derived from other sources. This point is addressed in the following section.

4. DEALING WITH INCOMPLETE DATA SETS

The discussion of the previous section illustrated that, ideally, it would be desirable to have a set of data that combines information on consumption, willingness to pay and socio-economic characteristics. At the same time, it indicated that there are very few sources of information that encompass all three of these aspects. Consumption data is most readily obtainable from water company client databases, and these lack any other kind of information. The most complete socio-economic information is likely to come from household surveys, which often lack data on water consumption. While willingness-to-pay data will often come from contingent valuation surveys, which may, or may not, collect adequate information on the other variables of interest.

In practice, the analyst will often have access to a reasonably good source of information on socio-economic data, while information on water consumption will either be missing or incomplete. As an illustration of how creative data analysis can be used to overcome the problems posed by incomplete data sets, this section presents some techniques that were developed in Panama with a view to inferring the following:

- unmeasured consumption;
- consumption from expenditure data;
- consumption from socio-economic characteristics.

These examples are intended to be illustrative rather than exhaustive. Other researchers will probably create new approaches that will enrich the practice in this field.

Inferring Unmeasured Consumption

The consumption of non-metered households cannot be directly observed, yet it is often relevant for subsidy design, especially when metering coverage is expected to increase in the near future. It is not generally reasonable to suppose that unmeasured households consume the same amount as measured households, for two reasons. First, the former face a marginal price of zero for water consumption and (other things being equal) are therefore likely to consume more than the latter, who face a positive price at the margin. Second, the distribution of meters across the population is not generally random, but typically shows a positive correlation with income levels and socio-economic status. If consumption behavior is also positively correlated with these variables, then there will be a selectivity bias when unmeasured consumption is inferred from the behavior of measured clients.

One possibility is to estimate unmeasured consumption indirectly by examining the behavior of a client during the first few months after a meter is first installed. It is reasonable to conjecture that previously unmeasured households will adjust their consumption gradually over time, or at least only after they have received their first measured bill. Therefore, water consumption during the first few months after a meter is installed may be informative as to the consumption level of unmeasured households, provided that the meter installation programme is relatively random with respect to the type of new households that are metered.

Table 4.1 presents monthly household consumption data for the first four months after the installation of a meter for the case of Panama. These households were identified from the IDAAN client database as those that, during the period of data available (May 1997 to April 1998), show a clear pattern of transition from unmeasured to measured bills. For example, the monthly data for a particular household shows that, during the first four months of the period, the water bill was based on unmeasured imputed consumption. Subsequently, every other bill is based on measured consumption. On this basis, it was inferred that the household in question had a meter installed during the fourth month.

The table shows that the average monthly consumption for households facing the residential tariff declined by about 25% over the first four months of metering (falling from 14,700 to 11,400 gallons per month)¹⁸. However, no such decline can be observed for the customers in the other two tariff bands, whose consumption actually increases slightly in the first few months following the introduction of metering.

Table 4.1: Average monthly consumption of newly measured clients between May 1997 and April 1998, Panama

Month since metering began		Type of user		
		Residential	Special	Interior
1 ^o month	Average ('000s gallons)	14,740	7,210	9,630
	No. of cases	152	799	2,171
2 ^o month	Average ('000s gallons)	10,020	7,790	10,620
	No. of cases	138	626	1,712
3 ^o month	Average ('000s gallons)	10,540	8,080	10,190
	No. of cases	37	607	1,901
4 ^o month	Average ('000s gallons)	11,400	7,650	10,770
	No. of cases	21	578	1,925

Note: The tariff structure included a minimum consumption level for each type of customer. For residential and interior customers, the minimum consumption level was 8,000 gallons per month. For special customers, it was 6,000 gallons per month. Presumably the registered consumption level is truncated at this minimum level in the client data set. In fact, bunching at these minimum levels can be observed in the data.

Source: Client database, IDAAN.

Another source of similar information comes from a metering programme conducted in a low-income neighborhood, known as Tocumen, in 1992. In that case, newly metered households were found to be registering a consumption of 18,900 gallons per month. Within 12 months of meter installation, however, this consumption fell to 12,000 gallons per month,

¹⁸ It is interesting to note that the final consumption level lies below the average for existing measured customers in this tariff band, which is 13,700 gallons per month. This may reflect the fact that initial metering efforts were directed towards households with relatively high consumption levels, in order to maximize their impact in restraining demand, while the last households to be metered have below average consumption. To rationalize the above results, it would have to assume that the new clients metered in 1997 to 1998 are in this last category.

which is close to the average consumption for measured households. These results suggest that unmeasured consumption may be as high as 66% above measured consumption. The differences in results between these two estimations may reflect differences in demand conditions between the two populations examined.

Finally, it is also interesting to compare these estimates of unmeasured consumption with the official estimates used by the water company to calculate the bills payable by unmeasured households. The summary statistics, which appear in Table 4.2 below, can be compared with the figures presented above. The average consumption imputed to unmeasured customers charged according to the special and interior tariff structures is remarkably close to that observed when these customers were initially metered. However, the estimated consumption for residential customers is substantially lower than the consumption levels observed in the first month of metering. The implication is that unmeasured residential customers are probably being undercharged in the present tariff structure.

Table 4.2: Mean and median of estimated average monthly consumption of unmeasured customers over the period April 1997 to April 1998

	Type of customer			
	Overall	Residential	Special	Interior
Mean	9.1	11.7	7.9	9.5
Median	7.0	9.1	6.0	8.2

Source: Client database, IDAAN.

Inferring Consumption from Expenditure Data

As noted above, the Living Standards Measurement Survey for Panama did not include direct information on water consumption, but only on water expenditure. Moreover, the water expenditure data collected was deficient in a number of respects. The main problems were the absence of information as to whether customers were measured or unmeasured, and on the tariff structure against which each customer was paying. Furthermore, it was not always clear whether the reported expenditure included or excluded the costs of refuse collection services that are jointly billed with water.

In order to gauge how substantial the divergence might be between actual water expenditure and that reported in the Living Standards Measurement Survey, histograms were plotted comparing the frequency distribution of expenditure in the survey as against the client database of the water utility, IDAAN. The resulting distributions for the standard residential tariff and the special social tariff are presented in Figures 4.1 and 4.2 respectively.

In both cases, there is a striking contrast between the two distributions. The key differences are as follows.

- The distribution of expenditure from the client database shows a marked concentration of households around the B./4.26, B./4.60, B./5.68 and B./ 8.00 (US\$ 4.26 TO US\$ 8.00) mark, which represents the minimum charge payable (depending

on the specific tariff structure). Moreover, the vast majority of clients seem to have bills around this level. Neither of these features is found in the distribution of expenditure from the Living Standards Measurement Survey, which presents a much flatter distribution of expenditure.

- Furthermore, in the first figure, there is a slight spike in the distribution at around the B./12 (US\$12) mark. Interestingly, this is exactly the level of expenditure that a residential household would incur if refuse collection charges were not subtracted from the water bill (B./6.40 for minimum water consumption on the standard tariff, and B./5.60 for refuse collection). This is suggestive of mistaken inclusion of the refuse collection charge in a significant number of cases.

Figure 4.1: Monthly water bill distribution from different information sources, standard residential customer, Panama

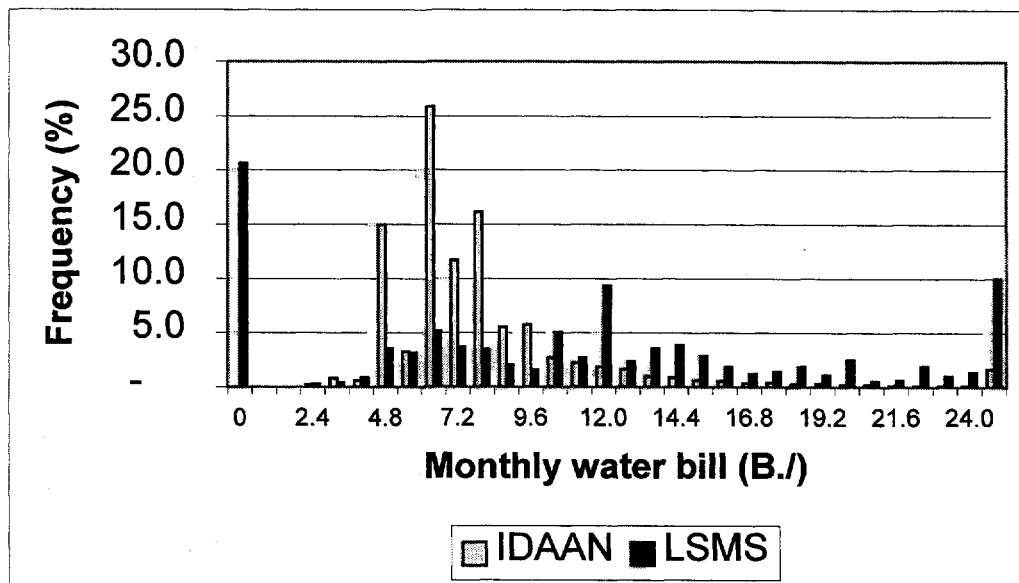
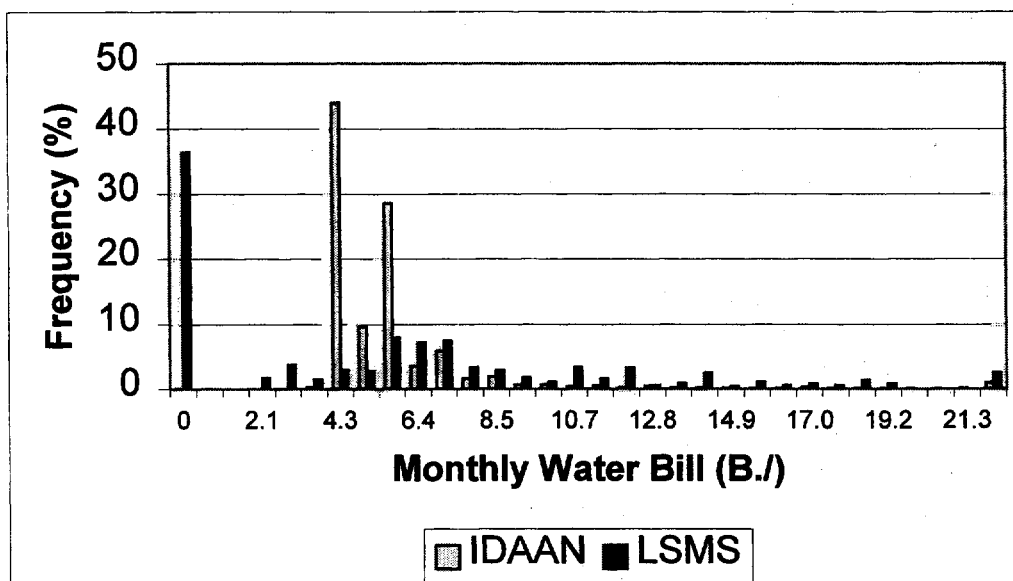


Figure 4.2: Monthly water bill distribution from different information sources, special residential customers, Panama



Notwithstanding these deficiencies, some attempt was made to retrieve consumption information from the expenditure data contained in the Living Standards Measurement Survey. In order to do this, the following assumptions had to be applied:

- all the reported expenditures related to water and sanitation services, and not to refuse collection services;
- households in metropolitan areas paid the metropolitan tariff, while those in the interior paid the provincial tariff;
- households in the first four deciles of the income distribution were paying in accordance with the concessional tariff structure;
- the water utility was correctly imputing the consumption of unmeasured households when calculating their bill.

On this basis, physical water consumption could be estimated for each household (albeit imperfectly) by applying the corresponding tariff structure to the water expenditure figure.

Inferring Consumption from Socio-Economic Data

Having performed the conversion from expenditure data to consumption data described above, a further problem remained with the Panama data. About 17% of households included in the Living Standards Measurement Survey stated that they did not pay their water bills and, hence, their reported expenditure was equal to zero. Clearly, in this case, the expenditure approach could not be applied and some alternative method would be required.

A suitable method for dealing with this kind of situation has been developed by Rajah and Smith (1993) in their work on the distributional impacts of introducing water metering in the UK. These authors had access to two data sets:

- a large-scale nationally representative household survey with a very rich set of information on income and other socio-economic indicators, but scant information about water consumption;
- a small-scale detailed local study of water consumption with a very limited range of socio-economic variables.

With the aim of incorporating a consumption variable into the household survey, Rajah and Smith (1993) estimate a water consumption function from the second data set based on socio-economic variables that are common to both data sets.¹⁹ The regression equation is stated formally below, where c is consumption, X are the common socio-economic variables and e is a stochastic error term. On the basis of this equation, it was then possible to predict consumption levels in the large-scale household survey by inserting the values of the corresponding socio-economic variables.

$$c = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k + e$$

This approach is clearly relevant to the estimation of the missing consumption variables in the Panama Living Standards Measurement Survey. Taking the inferred consumption levels from that part of the sample which reported water expenditure, a consumption function was estimated on the basis of the numerous socio-economic variables contained in the survey²⁰. These included household size, household expenditure, water service variables, geographical variables and a range of characteristics of the dwelling. The results, which are reported in Table 4.3, indicate that most of these variables are statistically significant in determining water consumption and that, overall, they explain some 37% of the variation in water consumption between households.

The consumption function was then used to assign an estimated consumption level to all households in the sample—*both* those with no recorded expenditure and those who reported positive expenditures on the water service. This ensured that the consumption of all households was estimated on a consistent basis. Table 4.4 reports some summary statistics for the estimated consumption levels. It is interesting to note that the average consumption levels for households below and above the poverty line are close to those observed in the IDAAN client database, although slightly larger.

Before leaving this subject, it is important to point out a number of limitations with this regression-based technique for imputing missing values from complementary data sets.

¹⁹ Arrellano and Meghir (1991) give a more technical treatment of the use of multiple data sets for estimation. See also Green (1997, Chapter 9) for the related topic of missing information in data sets.

²⁰ This approach may be subject to selectivity bias if non-reporting is correlated to the explanatory socio-economic variables on the right hand side of the equation.

- First, there are general econometric issues that arise with the above approach. For example, the use of only those explanatory variables that are available in both data sets may create an omitted variable problem in the first stage estimation of the consumption equation. Thus, the prediction of water consumption in the second stage may be biased.
- More fundamentally, the prediction of water consumption in the second stage will be the expected value of consumption conditional on the explanatory variables. These values will tend to be less dispersed than the original consumption data, so that extremely high or low consumption values disappear. The above problem may become important in settings where the treatment of the extreme values is a particularly sensitive policy issue. A relevant example would be the case of poor households with many family members, or with special medical needs, which require relatively high unavoidable water usage.

Table 4.3: Water consumption function estimated from the Panamanian Living Standards Measurement Survey

	Coefficient	Standard error	T-statistic
Dependent variable			
Monthly consumption (inferred) ('000s gallons)			
Independent variables			
Constant	2.7863	1.0156	2.74
Household size	0.0213	0.0062	3.43
Logarithm (household expenditure)	-0.4123	0.2370	-1.74
[Logarithm (household expenditure)] ²	0.0323	0.0139	2.32
Connected to sewer	0.2324	0.0272	8.55
Hours per day of water service	0.0115	0.0022	5.17
Owns a hose	0.0173	0.0412	0.42
Resides in an urban area	0.3563	0.0303	11.76
Resides in the metropolitan area	0.3298	0.0510	6.47
Reside in the province of Chiriqui	0.0351	0.0374	0.94
Owner occupier	0.0151	0.0517	0.29
Precarious housing	-0.4159	0.1283	-3.24
Rented housing	0.0669	0.0725	-2.72
Resides in a house	-0.2608	0.0959	-2.71
Resides in a shack	-0.6690	0.1437	-4.66
Resides in an apartment	-0.2905	0.1006	-2.89
Number of rooms	0.0385	0.0095	4.07
Number of bathrooms	0.0761	0.0230	3.30
Pays the residential tariff (inferred)	-0.1576	0.0510	-3.09
Number of observations	2045		
Adjusted R ²	0.3714		

Table 4.4: Mean and median monthly household consumption estimated from Living Standards Measurement Survey

Group	Units	Mean	Median
Below poverty line	gallons per month ('000s)	8.57	7.77
Above poverty line	gallons per month ('000s)	13.41	12.96

5. MODELLING AND SIMULATION ISSUES

The discussion above has focused on data requirements and sources, as well as a number of techniques for overcoming the problems posed by incomplete data sets. This section turns to the question of what needs to be done with the data once it has been collected, so that it can be used to answer the main design questions raised by a subsidy scheme.

All of the calculations described below can easily be done in a spreadsheet programme or other software. The creation of a simulation model to undertake these calculations should be regarded as a key stage in any subsidy design process. Specifically, there are three aspects of direct subsidy design where quantitative modelling can be particularly helpful:

- determining eligibility criteria for a subsidy scheme;
- establishing the appropriate magnitude for the subsidy;
- estimating the budgetary requirements of the subsidy programme.

Throughout the following discussion, it is important to note that the results of any quantitative analysis based on the sample of data collected for the purpose needs to be extrapolated to the population as a whole before any firm policy conclusions can be drawn. In the case of most household surveys, this is relatively straightforward, as the data sets typically include expansion factors. These are coefficients that reflect the number of households in the overall population that are represented in the survey sample by any particular observation.

Determining Eligibility Criteria

One of the principal reasons for merging all the above information into a single data set is to be able to study the errors of inclusion and errors of exclusion of different targeting mechanisms. The intention should always be to target water subsidies on the poor, defined as those with equivalent income below some specific threshold. However, in practice, household income is difficult to measure reliably. Eligibility for the subsidy would therefore typically be determined with reference to some other more readily observable variable(s) that has a demonstrated correlation with poverty.

Experience shows that it is difficult to find good proxy variables for poverty, and the suitability of any candidate variable should therefore be rigorously assessed. It is presumed that the data set constructed will contain information both on household income and on the values of a range of possible proxy variables. Using this information, a simulation model can be developed to test for the targeting properties of alternative proxies. The model should be able to perform the following tasks.

- Identify how many households in the sample are in the target population, in the sense that they are genuinely poor. The target population would correspond to all households for which the poverty indicator variable $P \leq P^*$. For example, P could represent household equivalent income, and P^* the first quintile level of this variable.

In this case, the target population corresponds to the 20% of households with the lowest equivalent income.

- Identify how many households in the sample would be identified as eligible for the subsidy on the basis of any particular proxy variable(s). The eligible population would correspond to all those with socio-economic characteristics, Z , complied with the eligibility criterion $Z \leq Z^*$. For example, Z could represent the presence of a telephone connection in the household. In this case, the eligible population corresponds to those who do not have a household telephone connection.
- Calculate the proportion of households in the target population who would not be deemed eligible in terms of the proxy variable; these are the errors of exclusion (E_E). For example, it may be that a few households in the first income quintile have a telephone connection for some exceptional reason. More formally, these can be expressed as follows.

$$E_E = \frac{\sum_{i=1}^N I_i(Z_i > Z^*; P_i \leq P^*)}{\sum_{i=1}^N I_i(P_i \leq P^*)}$$

- Calculate the proportion of households considered eligible who would not belong to the target population; these are the errors of inclusion (E_I). For example, it may be that many of the households which lack a telephone connection have income levels above the first quintile of the distribution. More formally, this can be expressed as follows.

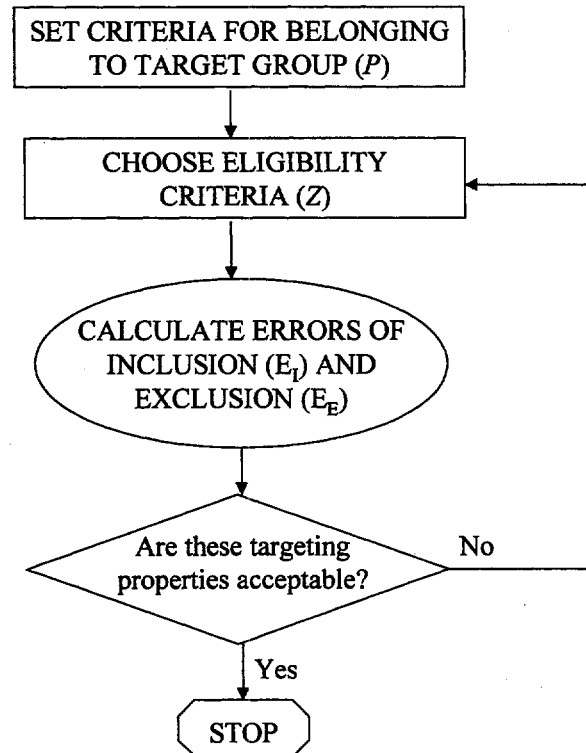
$$E_I = \frac{\sum_{i=1}^N I_i(Z_i \leq Z^*; P_i > P^*)}{\sum_{i=1}^N I_i(Z_i \leq Z^*)}$$

On the basis of this analysis, the simulation model should be able to produce output along the lines illustrated in Table 5.1. By varying the eligibility criterion, the policymaker should be able to compare the targeting properties of alternative proxy variables and iteratively establish the most suitable targeting criteria, along the lines illustrated in Figure 5.1.

In the case of Panama, this type of analysis was critical in leading to the rejection of a geographically based criterion for the water subsidy and in identifying which household level variables should be considered as a basis for eligibility.

Table 5.1: Format of output from simulation model for analysis of eligibility criteria

Target population	First income quintile		
Eligibility criterion	Absence of telephone connection		
Results		Meets eligibility criterion	Fails to meet eligibility criterion
Member of target population		A	B
Non-member of target population		C	D
Errors of exclusion	$B/(A+B)$		
Errors of inclusion	$C/(A+C)$		

Figure 5.1: Iterative process for establishing the most suitable targeting criteria

Establishing the Magnitude of Subsidy

As stated earlier, a reasonable approach to setting the level of the subsidy is given by the following formula.

$$S = \begin{cases} T(c) - WTP(c) & \text{if } T(c) - WTP(c) > 0 \\ 0 & \text{if } T(c) - WTP(c) \leq 0 \end{cases}$$

In order to apply this formula it is necessary to have information on the following two variables, both of which should appear in the overall data set.

- The first is the consumption level of each household, in order to be able to work out the bill that would apply given the applicable tariff structure.
- The second is the income level of each household, given that willingness to pay is often expressed as a percentage of household income.

Ideally, it would be optimal to give each individual household a subsidy level consistent with the above formula, using the consumption and income levels corresponding to each household. However, such an individually tailored subsidy would be extremely expensive to administer. Therefore, in practice it is necessary to calculate the required magnitude of the subsidy with respect to some sort of average or representative household from the target group.

A first approach would be to use the average values of the target population for each variable²¹. That is, to set the subsidy level as the difference between the average bill and the average willingness to pay of the target population. This subsidy level would then be applied to all eligible households. However, within the target group, there will be some households that are relatively better off than others, and some that have relatively high essential needs compared to others. For some of these households, the subsidy will be insufficient to induce them to consume the required amount of water, while, for others, it will be excessive. It is therefore important to check the effects of this subsidy by modelling the impact of the 'average' subsidy on each individual household. This can be estimated as follows.

$$d_i = (WTP_i + S - T(c_i)) \times I_i(P_i \leq P_i^*)$$

If d is very negative for a substantial number of households, then using an average value for estimating the subsidy level may not be appropriate. This, in turn, will depend on the dispersion in consumption and willingness to pay within the target population group.

In the above case, perhaps it would be advisable to explore the effects of setting the subsidy level according to the values of consumption and willingness to pay of, say, the poorest 25% of the target population. This approach will almost certainly increase the level of subsidy, thereby offering more protection for the poorest of the poor. However, this comes at the expense of increasing benefits to the other poor households that do not really require this level of support, and thereby raises the budgetary requirements of the subsidy. A second alternative is to explore the possibility of a differentiated subsidy, one level for very poor households and another (lower) for the relatively better off among the target population. This approach would save budgetary resources, but it increases the administrative costs of the programme because it now becomes necessary to identify two separate categories of consumers.

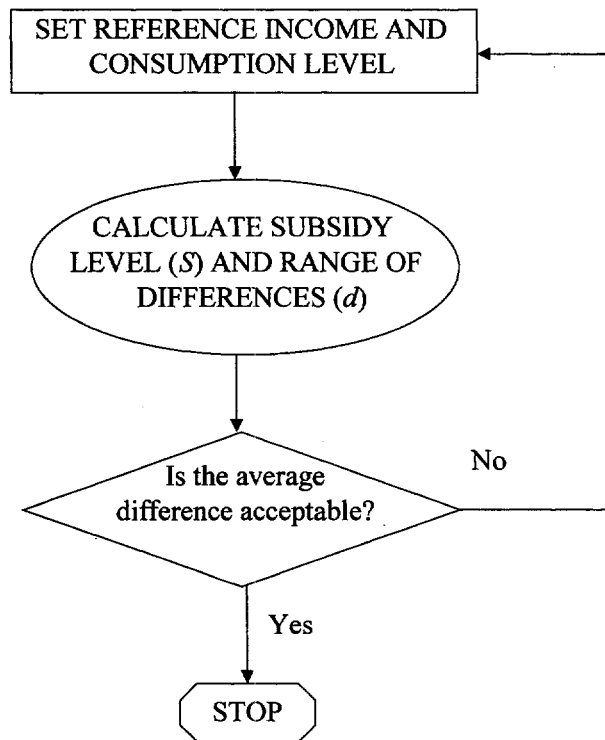
²¹ Average values should be calculated using the expansion factors of each observation in the sample as weights.

An approach of this sort was used in Chile where subsidies are set regionally in order to account for varying income and tariff levels across the country.

By varying the consumption and income reference levels, the model should allow the policymaker to determine iteratively the appropriate magnitude of the subsidy, along the lines illustrated in Figure 5.2.

The simulations undertaken for the case of Panama showed that, while most poor households did not require a subsidy for water consumption, the poorest among the poor would require one. Based on these results, it was recommended that only a very small fraction of households (those in extreme poverty) should benefit from a water consumption subsidy.

Figure 5.2: Iterative process for determining the appropriate magnitude of the subsidy



Estimating Budgetary Requirements

Another element that should inform the design of a water subsidy scheme is its total budgetary requirement, as well as the related administrative costs. The direct costs of the subsidy are easily estimated as the number of eligible households times the value of each subsidy. (If there were some differentiation of the subsidy level by household, then the

formula below should be adjusted to take this into consideration.) The basic formula is as follows.

$$B = \sum_{i=1}^N S \cdot I_i(P_i \leq P^*) \times \delta \times \rho$$

Note the presence of the multipliers δ and ρ . The first of these, δ , represents the rate of uptake of the subsidy, since not all the households in the target population will apply to receive the subsidy. The lower the percentage of households in the target group that apply, the smaller the budgetary requirements. The second of these parameters, ρ , represents the rate of infiltration. This is because, however rigorous the screening procedures, it is likely that some households outside of the target group will succeed in demonstrating their eligibility. The higher the rate of infiltration, the larger the budgetary requirements.

The parameters ρ and δ cannot be predicted with precision. Rather, it is best to experiment with a range of alternative assumptions in order to examine the sensitivity of the budgetary requirement to variations in these coefficients.

The formula given above provides a prediction for any given year. For fiscal planning, it may be convenient to project expenditure on the programme for future years. This is relatively simple to do, provided that some dynamic assumptions are made with respect to the growth in the number of eligible households.

To estimate the administrative costs of a subsidy programme requires more effort. They will include the costs of processing the applications for the subsidy, administering household interviews, and managing the information system needed to keep track of eligible households. A particularly important performance parameter for a subsidy programme is the ratio of administrative costs to the overall costs of the subsidy.

In the case of Panama, administrative costs were estimated based on a series of assumptions regarding:

- the number of households which were eligible for the subsidy;
- the frequency with which eligibility should be reassessed (Y);
- the rate of uptake (δ) and of infiltration (ρ) which, together, determine the number of interviews to be undertaken for each eligible household identified;
- the mark-up of fixed administrative costs over variable administrative costs (ϕ);
- the number of interviews that a social worker can undertake in a year (V);
- the salary of each social worker (w).

The general formula is stated below.

$$A = \left(\frac{\sum_{i=1}^N I_i (P_i \leq P^*)}{Y} \right) \times \delta \times \rho \times \varphi \times V \times w$$

Some of the parameters for this estimation were taken from the previous experience of social workers in the existing public-sector water utility, IDAAN. Considerable sensitivity analysis was undertaken with respect to the key parameters.

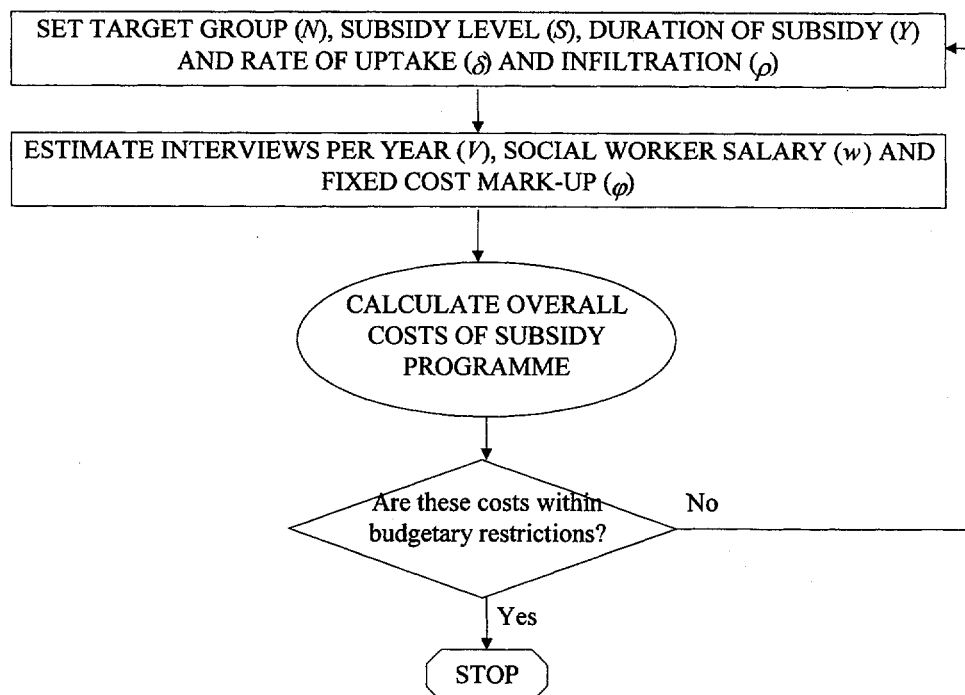
In parenthesis, it should be noted that the administrative costs of a water subsidy programme are lower when there is already a targeting mechanism in place in the country that is used to allocate other welfare benefits. In this case, only the additional or marginal administrative cost of the subsidy should be attributed to the water programme.

On the basis of this analysis, the simulation model should be able to produce output along the lines illustrated in Table 5.2. By varying the assumptions made about the key parameters, the policymaker should be able to determine the overall costs of the programme, and to establish iteratively the subsidy design parameters that are compatible with the overall budgetary, along the lines illustrated in Figure 5.3.

Table 5.2: Format of output from simulation model for analysis of financial and administrative costs

Subsidy level	S	Duration of subsidy	Y
Target group	N	Target group	N
Rate of uptake	δ	Interviews per eligible household	$\delta \times \rho$
Rate of infiltration	ρ	Number of interviews per year	V
Population growth rate	g	Salary of social worker	w
		Fixed-cost mark-up	ϕ
Direct subsidy costs (D)	$S \times N \times \delta \times \rho \times (1+g)^t$	Indirect administrative costs (I)	$(N/Y) \times \delta \times \rho \times V \times w \times \phi$
		Performance parameter	$I/(D+I)$

Figure 5.3: Iterative process for establishing an affordable set of subsidy parameters



6. CONCLUSIONS

This paper has discussed some of the practical issues that have to be dealt with when designing a rational water subsidy scheme. A central conclusion is that the choice of suitable design parameters for a subsidy scheme needs to be supported by appropriate empirical analysis to simulate the impact of alternative types of subsidies on the ultimate target population. Without such investigation, there is little guarantee that a subsidy system, however well meant, will meet its intended objectives.

Nevertheless, it is acknowledged that the kind of analysis required to design a subsidy programme is relatively demanding in terms of information. Ideally, the policymaker would need to have access to a single and consistent data set containing information on willingness to pay, water consumption and a range of socio-economic characteristics, each reported at the household level.

Clearly, such comprehensive information will rarely be available in practice. However, this in itself should not be regarded as a sufficient reason to abandon the counsels given above. As has been illustrated in the paper, deficiencies in the available information can to some extent be overcome by collating different sources of data and imaginatively manipulating this data so as to generate estimates of the missing variables.

Based on the review conducted above, the most valuable sources of information are likely to be the following.

- *Water company client databases*—these provide robust information on the measured consumption of formal customers, but provide little information about unmeasured consumption or informal customers. The main drawback with this data source is that it will not contain any information on the socio-economic characteristics of the clients or their willingness to pay.
- *General socio-economic household surveys*—most countries conduct extensive household surveys on an occasional basis, many of them modelled on the World Bank's 'Living Standards Measurement Survey' blueprint. Such surveys provide excellent socio-economic data, but tend to record water expenditure rather than water consumption. Various deficiencies in the recording of water expenditure can make it difficult to make reliable inferences about water consumption.
- *Willingness-to-pay surveys*—these are generally tailor-made surveys conducted in the context of a specific project. They are an extremely flexible tool for determining willingness to pay, and may also be used to complete other gaps in the information base. However, they are subject to a range of methodological criticisms. Perhaps the most fundamental of these is the fact that they are based on hypothetical as opposed to real behavior. Where it is not possible to conduct a willingness-to-pay survey, international benchmark values may be used, based on previous experience.

Once a suitable data set has been constructed from these sources, a simulation model can be created using simple spreadsheet software. The model should be capable of addressing the following key design questions.

- *What should the eligibility criteria be?* Subsidies are generally assigned on the basis of eligibility criteria presumed to have a strong correlation with poverty. Using the subsidy data set, the targeting performance of alternative eligibility criteria can be compared by examining the proportion of the poor who would meet any particular eligibility criterion. It is also possible to examine what proportion of the eligible population are not genuinely poor, for any particular set of criteria.
- *How large should the subsidy be?* Economic theory suggests that the subsidy should be set equal to the difference between the willingness to pay of the poor and the cost of purchasing a subsistence level of consumption. However, willingness to pay is likely to vary with household income and subsistence consumption with household composition. Using the simulation model, it is possible to experiment with alternative reference levels of income and consumption, and to examine what proportion of poor households would be adequately protected by any given level of the subsidy.
- *How much will the subsidy scheme cost?* The costs of the subsidy programme must necessarily be commensurate with the available funding. A simulation model allows ready estimation of the overall costs of the subsidy programme. These depend not only on the size and scope of the subsidy, but also on the rate of uptake among eligible households, as well as the rate of infiltration of non-eligible households. The administrative costs of the programme also need to be taken into account, and, in particular, the proportion of total costs that are absorbed by administrative procedures.

Armed with this information, the policymaker should be in a position to design a subsidy programme that reaches the intended beneficiaries, provides them with the level of financial support which is strictly necessary, meets the overall budgetary restrictions, and does not waste an excessive amount of funding on administrative costs.

REFERENCES

- Ardila, S., Quiroga, R., and Vaughan, W.J. (1998), 'A Review of the Use of Contingent Valuation Methods in Project Analysis at the Inter-American Development Bank', paper presented at the National Science Foundation Workshop, *Alternatives to Traditional Contingent Valuation Methods in Environmental Valuation*, Vanderbilt University, Nashville, Tennessee, October 15th–16th 1998.
- Arrellano, M., and Meghir, C. (1991), 'Labour Supply and On-the-Job Search: An Empirical Study Using Complementary Data Sets', *Review of Economic Studies*.
- Carpentier, A., and Vermersch, D. (1997), 'Measuring Willingness to Pay for Drinking Water Quality Using the Econometrics of Equivalent Scales', *Nota Di Lavoro* 92.97, Fondazione Eni Enrico Mattei.
- Council for the Protection of Rural England, Environmental Agency, Office of Water Services, Royal Society for the Protection of Birds, Water Services Association, UK Water Industry Research Limited (1998), 'Towards an Environmentally Effective and Socially Acceptable Strategy for Water Metering in the UK', research paper produced by OXERA, Oxford.
- Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, published for the World Bank by John Hopkins University Press, Baltimore and London.
- Diamond, P.A., and Hausman, J.A. (1994), 'Contingent Valuation: Is Some Number Better than No Number?' *Journal of Economic Perspectives*, 8:4, 45–64.
- Foster, V., Bateman, I.J., and Harley, D. (1997), 'Real and Hypothetical Willingness to Pay for Environmental Preservation: A Non-Experimental Comparison', *Journal of Agricultural Economics*, 48:2, 123–38.
- Freeman III, A.M. (1994), 'The Measurement of Environmental and Resource Values: Theories and Methods', *Resources for the Future*, Washington DC.
- Green, W.H. (1997), *Econometric Analysis*, Third Edition, Prentice-Hall International, Inc.
- Hansen, L.G. (1996), 'Water and energy price impacts on residential water demand in Copenhagen', *Land Economics*, 72(1), pp. 66-79.
- Hewitt, J.A., and Hanemann, W.M. (1995), 'A Discrete/Continuous Choice Approach to Residential Water Demand under Block Rate Pricing', *Land Economics*, 71:2, 173–92.
- Lyman, R.A. (1992), 'Peak and Off-peak Residential Water Demand', *Water Resources Research*, 28:9, 159–67.
- National Oceanographic and Atmospheric Administration (1994), 'Natural Resource Damage Assessment: Proposed Rules', *Federal Register*, 59:5, 1062–91.

Nieswiadomy, M.L., and Molina, D.J. (1989), 'Comparing Residential Water Demand Estimates under Decreasing and Increasing Block Rates Using Household Data', *Land Economics*, **65**:3, 280–9.

Rajah, N., and Smith, S. (1993), 'Distributional Aspects of Household Water Charges', *Fiscal Studies*, **14**:3, 86–108.

Stephens, T.H., Miller, J., and Willis, C. (1992), 'Effect of Price Structure on Residential Water Demand', *Water Resources Bulletin*, **28**:4, 681–5.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS2328	Social Transfers and Social Assistance: An Empirical Analysis Using Latvian Household Survey Data	Branko Milanovic	April 2000	P. Sader 33902
WPS2329	Improving Russia's Policy on Foreign Direct Investment	Joel Bergsman Harry G. Broadman Vladimir Drebenstov	May 2000	S. Craig 33160
WPS2330	Reducing Structural Dominance and Entry Barriers in Russian Industry	Harry G. Broadman	May 2000	S. Craig 33160
WPS2331	Competition, Corporate Governance, and Regulation in Central Asia: Uzbekistan's Structural Reform Challenges	Harry G. Broadman	May 2000	S. Craig 33160
WPS2332	Financial Intermediary Distress in the Republic of Korea: Small Is Beautiful?	Paola Bongini Giovanni Ferri Tae Soo Kang	May 2000	K. Labrie 31001
WPS2333	Output Fluctuations in Latin America: What Explains the Recent Slowdown?	Santiago Herrera Guillermo Perry Neile Quintero	May 2000	C. Palarca 35328
WPS2334	Sex Workers and the Cost of Safe Sex: The Compensating Differential for Condom Use in Calcutta	Vijayendra Rao Indrani Gupta Smarajit Jana	May 2000	P. Sader 33902
WPS2335	Inflation and the Poor	William Easterly Stanley Fischer	May 2000	K. Labrie 31001
WPS2336	Endogenous Enforcement and Effectiveness of China's Pollution Levy System	Hua Wang David Wheeler	May 2000	Y. D'Souza 31449
WPS2337	Pollution Charges, Community Pressure, and Abatement Cost of Industrial Pollution in China	Hua Wang	May 2000	Y. D'Souza 31449
WPS2338	The Geography of International Investment	Howard J. Shatz Anthony J. Venables	May 2000	L. Tabada 36896
WPS2339	Building Subnational Debt Markets in Developing and Transition Economies: A Framework for Analysis, Policy Reform, and Assistance Strategy	Michel Noel	May 2000	M. Noel 32581

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS2340	Currency Substitution in Latin America: Lessons from the 1990s	Pere Gomis-Porqueras Carlos Serrano Alejandro Somuano	May 2000	M. Puentes 39621
WPS2341	The Tyranny of Concepts: CUDIE (Cumulated, Depreciated Investment Effort) Is <i>Not</i> Capital	Lant Pritchett	May 2000	R. Widuri
WPS2342	What Can We Learn about Country Performance from Conditional Comparisons across Countries?	Martin Ravallion	May 2000	P. Sader 33902
WPS2343	Ownership and Performance of Lithuanian Enterprises	David A. Grigorian	May 2000	D. Brown 33542
WPS2344	Designing Direct Subsidies for Water and Sanitation Services: Panama—A Case Study	Vivien Foster Andrés Gómez-Lobo Jonathan Halpern	May 2000	S. Delgado 37840